14.771 Problem Set 1

1 Gender Bias in Investment in Children: Theory and Regressions

Two classical related questions economists have asked are (i) whether parents have gender preferences in fertility, and (ii) whether parents invest more in boys than in girls (e.g., Pande and Astone (2007), Jayachandran and Kuziemko (2011)). In this section, we will have a brief discussion of this classic issue in development.

Exercise 1

Assume that families have one child, and they have to make a one-time investment in the children's education. The parent's utility is given by

$$u(h,c) = h^{\alpha} c^{\beta} \tag{1}$$

with budget-constraint

$$ph + c \leq I \tag{2}$$

where h is human capital, c is consumption, $p > 0$ is the unitary price of human capital, consumption's price is normalized to 1, and $I > 0$ is the family's income. The parameters α and β captures the parent's preferences towards investment in children's human capital and family's consumption.

1. What is the human capital investment chosen by a family maximising (1) subject to (2)?
2. Assume now that the price of human capital may differ by gender, where p_G is the price for girls and p_B for boys. Cite two reasons why the price of education for boys and girls might differ (max. 3 lines per reason) (tip: you can define price broadly, not only school tuition, for example).
3. Suppose that in equilibrium, the investment for a girl is smaller than the investment for a boy in a given family. Under this model, is this irrefutable evidence of parental gender *preferences* regarding their children education? Explain your answer in at most 3 lines.
4. Assume families have exactly two children, that their utility is given by $u(h_1, h_2, c) = h_1^{\alpha} h_2^{\alpha} c^{\beta}$, and that the budget constraint is $ph_1 + ph_2 + c \leq I$. Will the average investment in human capital *per* child fall in comparison to the family with a single child?

Exercise 2

Suppose in a society each family has a single child. You have access to a *cross-section* with information on each child's gender (dummy M_i, equal to 1 if male), and y_i, the child's education level. Answer the following questions, with maximum 3-4 sentences each.

1. (i) Write down which regression you would estimate to understand whether boys receive more parental investment, and (ii) the relationship between the independent variable and the error term necessary for identifying that this gap is investment is *caused* by gender.
2. Provide two reasons why you might worry that this this relationship may not hold in the setting described above.
3. Assume the equation you wrote on the item above indeed capture differences in parental investment *caused* by the child's gender. Can you interpret your results as a clear evidence of parental *preferences* towards a gender or another? What if you knew that $p_G = p_B$?
4. Even in the context of RCTs, researchers often control for covariates even when they are convinced the randomization worked. Why can this be a good idea?
5. Assume you observe a variable capturing the number of months the child was breastfed, which you verify to be positively correlated with the child's level of education. Do you think it is a good idea to control for this variable?

2 Big Push Graduation Program: Designing and Analysing RCTs

This section asks you to think about methodological issues in RCTs in the context of Banerjee et al. (2015). This paper will be covered in class on September 15, but – because the questions are primarily methodological – most of the questions can be answered prior to this.

Exercise 3 (max 3 sentences per question)

1. In three of the six countries, the authors used a two-step randomisation process: they first randomised villages into either treatment or control, and then within treatment villages they randomised households into either treated or untreated. This means there are three separate groups: (i) treated households in treatment villages, (ii) untreated households in treatment villages, (iii) control households in control villages. How does this design allow them to measure spillovers within a village? Be precise about which groups will be compared to which.
2. Are there any drawbacks to using this kind of design?

Exercise 4

Please download the dataset, *ps1_data.csv*, constructed from Banerjee et al. (2015). It consists of the consumption outcome variable for the three countries in which there was two-stage randomisation.

1. For now, we will ignore the control household in treated villages, and just focus on comparing treated households in treated villages to households in control villages. Write out an estimating equation to estimate the effect of being assigned to treatment on the consumption index.

2. Using the data, run the corresponding regression. Report and interpret the results.

3. Now, run a regression that yields a coefficient that can be interpreted as the strength of spillovers. Do you find evidence of important spillover effects?

Exercise 5

In this exercise, we will pay close attention to the meaning of standard errors and what are the "correct" standard errors to use in this setting. For the sake of simplicity, we will again ignore the untreated households in treatment villages – you can drop them from the dataset.

1. Describe what a standard error is and why we care about it (max 3 sentences).

2. A common rationale for why we need standard errors is that we observe only a finite sample of observations from a very large population, and we are trying to learn about a parameter (here, a treatment effect) for the entire population. Sometimes, we compute standard errors using analytical formulas (think of the typical "robust" or "clustered" standard errors).

   A different approach is to use the bootstrap. The bootstrap calculates standard errors by leveraging the above intuition for why we need standard errors, by repeatedly re-sampling "bootstrap" sub-samples from our main sample (treating it as a population). We then re-run the regression on each one, and observe the distribution of the estimated treatment effects. If the distribution has high variance, then this suggests that the process of sampling from a larger population involves considerable randomness and thus our standard errors should be large.

   We would like you to implement the bootstrap and plot the distribution of treatment effects. In particular, please do this through the following steps:

   (a) Define the number of villages in the dataset as N.

   (b) Create a new dataset by randomly drawing, *with* replacement, N of these villages. This is our bootstrap sub-sample, b.

   (c) Regress the consumption index on treatment. Call the coefficient β_b.

   (d) Repeats Steps (b) – (c) a total of B times, for $B = 800$.
   (e) Plot a histogram (or density plot) of $\{\beta_b\}_{b=1,2,...,B}$ and report the standard deviation, which serves as an estimate for the standard error.[1]
   (f) Compare this estimate to the typical "robust" (Eicker-Huber-White) standard error, obtained from regressing the consumption index on treatment in the original dataset (still dropping the untreated households in treatment villages). Do you have any intuition for why they might be different? Which (if any) rejects the null?

3. Now, re-run the same regression, but with standard errors clustered at the village level. Compare this standard error to the bootstrapped one.

4. It seems that the robust and clustered standard errors are meaningfully different. Sometimes, people suggest that you should use the clustered standard errors because outcomes are correlated within a cluster, and we need to account for that. However, that rationale alone makes it hard to see why we should cluster at the village level and not, for example, at the country level or perhaps at an industry level.

   One way to work out which approach is appropriate is to run a Monte Carlo exercise. We will simulate fake data in which the true treatment effect is zero, where a researcher analyses a randomised experiment that looks just like Banerjee et al. (2015). We will do this a large number of times, and see how often we (falsely) reject the null hypothesis of zero treatment effect at the 5% level using the different ways of computing standard errors. Whichever method is "correct" should reject the null about 5% of the time.

   In particular, please implement this with the following steps:

   (a) Create $G = 50$ villages and randomly assign half to treatment.
   (b) Create a village-level "shock" variable, $\varepsilon_v \sim N(0, 3)$.
   (c) For each village, create $N = 100$ households.
   (d) Create a household-level "shock" variable, $\varepsilon_h \sim N(0, 3)$.
   (e) Let the outcome variable we are interested in be given by $y = \varepsilon_v + \varepsilon_h$ (note that this implies that the treatment effect is zero).
   (f) Regress y on treatment using robust and clustered standard errors. Save the coefficient, and the two different standard errors. Call these $[\beta_r, SE1_r, SE2_r]$.
   (g) Repeat (a)-(f) a total of $R = 800$ times. Make sure you change the seed.
   (h) For each type of SE, compute and report the share of simulations that rejected the null hypothesis of zero treatment effect (i.e., $|\beta_r| > 1.96 \cdot SE_r$). Which one do you conclude is appropriate for our setting?

---

[1] This is sometimes referred to as the "pairs bootstrap-se" or "pairs bootstrap-c" method, because we resample the (y, T) as a pair, calculate the coefficient, and use these to obtain the standard error.

   (i) Finally, re-run this Monte Carlo exercise, but assign treatment at the individual level instead (rather than assigning whole villages to treatment). What do you find? Given that nothing has changed about the data generating process, what does this imply about the often-suggested rationale for clustering (i.e., that we need to cluster when there is correlation in the outcome variable within clusters)? For more on this, see Abadie, Athey, Imbens, and Wooldridge (2017).

5. Like most analytical standard errors, the "clustered" standard error estimator is only asymptotically valid, so when we use it in finite samples we justify it as an approximation. Moreover, it converges as G, the number of clusters, goes to infinity. This means that the asymptotic approximation may be poor if there are few clusters (even if there are very many observations within each cluster!).

   As a potential solution to this, researchers have pointed out that the bootstrap may perform better if G is small. This final question explores this. We will repeat the exercise of 5.4, but this time we will also calculate bootstrapped standard errors. We will make three changes to 5.4: (i) set $G = 10$, (ii) set the variance of the village-level shock to 6, (iii) bootstrap the t-statistic rather than the coefficient. Bootstrapping the t-statistic is a little more complicated, but has the same basic intuition. You can do it in the following way, for each of the $R = 800$ simulated datasets.

   (a) Run a regression of y on treatment using clustered standard errors. Call the t-statistic w and the coefficient β.

   (b) Create a bootstrap sub-sample (just as you did in 5.2), and run the same regression (with clustered standard errors). Calculate the t-statistic as $w_b = (\beta_b - \beta)/SE_b$.

   (c) Repeat (b) $B = 400$ times, and define the 2.5th and 97.5th percentiles of $\{w_b\}_b$ as $w_{[2.5]}$ and $w_{[97.5]}$.

   (d) Reject the null hypothesis if $w < w_{[2.5]}$ or $w > w_{[97.5]}$.

   Again, report the share of simulations in which you (incorrectly) reject the null hypothesis of zero treatment effect. Compare this to the clustered standard errors. Which one achieves closer to the nominal 5% size? For more discussion on this, see Cameron, Gelbach, and Miller (2008).

   Hint: For 5.5, you might want to run this remotely on the MIT Econ servers. See the recitation slides for some guidance on how to do this. If you don't have access, it is fine to do it locally (you can use smaller R or B if you like).

   Hint 2: In Stata, you can bootstrap the t-statistic with the command "bootstrap" (writing the expression in 5.5b), and the 2.5th and 97.5th percentiles of this expression will be stored in the matrix "e(ci_percentile)".

References

A. Abadie, S. Athey, G. W. Imbens, and J. Wooldridge. When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research, 2017.

A. Banerjee, E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236), 2015.

A. C. Cameron, J. B. Gelbach, and D. L. Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.

S. Jayachandran and I. Kuziemko. Why do mothers breastfeed girls less than boys? evidence and implications for child health in india. *The Quarterly journal of economics*, 126(3):1485–1538, 2011.

R. P. Pande and N. M. Astone. Explaining son preference in rural india: the independent role of structural versus individual factors. *Population Research and Policy Review*, 26(1):1–29, 2007.

14.771: Development Economics
Fall 2021